

For Reference

NOT TO BE TAKEN FROM THIS ROOM

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex libris
UNIVERSITATIS
ALBERTAEASIS





Digitized by the Internet Archive
in 2019 with funding from
University of Alberta Libraries

<https://archive.org/details/Small1967>

THE UNIVERSITY OF ALBERTA

THE USEFULNESS OF
CANADIAN ARMY SELECTION TESTS
IN A CULTURALLY RESTRICTED POPULATION

A DISSERTATION
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF ARTS

FACULTY OF ARTS
DEPARTMENT OF PSYCHOLOGY

by

GEORGE WALTER SMALL, B.A.

MONTREAL, P.Q.

APRIL, 1967

University of Alberta
Faculty of Arts
Department of Psychology

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "The Usefulness of Canadian Army Selection Tests in a Culturally Restricted Population", submitted by George Walter Small, B.A., in partial fulfilment of the requirements for the degree of Master of Arts.

SYNOPSIS

A study was made of the effectiveness or usefulness of two Canadian Army tests as pre-enrolment screening devices in a "culturally-restricted" population of applicants for service in the Canadian Forces. It was hypothesized that the tests would show a discriminatory bias against the group tested since the tests were standardized on non-restricted applicant populations. The screening devices used were the Select-R Test, which is non-verbal and designed to be relatively "culture-free", and the Select-A Test, which is highly loaded with verbal content.

The experimental sample consisted of male Indian, Eskimo and Metis youth, from the Northwest Territories, who had expressed interest in serving with the Canadian Forces.

Criteria selected for assessing the effectiveness of the tests in the study were (1) significant similarity of mean test scores of both experimental and standardization samples; (2) low correlation of test scores of the experimental sample with education and age; and (3) high correlation of the test scores of the sample with a dichotomous, subjective assessment of suitability for enrolment made by the testing officer. A detailed item analysis of both tests was also carried out to determine the items in the tests on which the experimental sample differed significantly from standardization samples in number of correct responses.

The hypothesis of bias was partially proven. The Select-A Test proved much more difficult for the experimental group and

correlated more highly with educational level than the Select-R which, in turn, correlated more highly with the subjective assessment of suitability. On the average, the experimental sample also achieved better results on the Select-R than the standardization groups. Both tests showed negative correlation with age, suggesting that older applicants are of generally lower quality, though the experimental sample was too small to permit any firm conclusions in this regard.

The item analysis revealed that no particular type of item on either tests was specifically easier or more difficult for the sample tested. However, on the whole, the experimental sample achieved as high a percentage of correct responses as did the standardization group on the Select-R but did much more poorly on the Select-A.

It was concluded that (1) the Select-R test is much less influenced by the cultural, educational and environmental background of the individual tested than is the highly-verbal Select-A test; (2) that revision of the Select-R test for use with the population from which the experimental sample was drawn is unnecessary; and (3) though both tests are useful as screening devices in the Far North, the non-verbal Select-R displays less bias towards the non-white, "culturally-restricted" applicant population and should be used with this group in preference to the Select-A Test.

On the basis of this study, the Select-A Test was replaced by the Select-R Test as the preliminary screening device for Canadian Forces applicants of non-white origin in the Northwest Territories and Yukon Territory.

ACKNOWLEDGEMENTS

I wish to extend thanks to Colonel WRN Blair and Major WJ Lott of the Canadian Army, who have provided the opportunity for this study and the consistent encouragement which enabled the project to be completed. Special thanks are tendered to Mrs Rhoda Cucheran, Miss Joyce Ling and Mlle Denise Lavigne, who typed the manuscript, and to Mr Norm Burgess of the Department of Indian Affairs in Yellowknife for his continued interest and support.

G.W.S.

TABLE OF CONTENTS

	<u>Page</u>
Title Page	i
Examiners Certificate	ii
Synopsis	iii
Acknowledgements	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
I INTRODUCTION	
General Considerations	1
The Problem	5
Definition of Terms Used	6
II SURVEY OF THE LITERATURE	
History of "Culture-Reduced" Tests	12
Development of the Select-R Test	14
Development of the Select-A Test	22
Studies on Culturally-Restricted Canadian Groups	27
III PROCEDURE	
Description of the Select-R Test	31
Description of the Select-A Test	32
Test Administration and Scoring	34
Selection of the Experimental Group	35
Composition of the Select-R Standardization Group	36
Composition of the Select-A Standardization Group	37
Method of Determining the Results	38
IV RESULTS	
<u>Part I</u>	
Select-R Scores and Comparison of Groups	39
Select-A Scores and Comparison of Groups	41
Intercorrelations and Correlations	46

Page 1

Example 1

Example 2

Example 3

Example 4

Example 5

Example 6

Example 7

Example 8

Example 9

Example 10

Example 11

Example 12

Example 13

Example 14

Example 15

Example 16

Example 17

Example 18

Example 19

Example 20

Example 21

Example 22

Example 23

Example 24

CHAPTER		<u>Page</u>
IV	RESULTS (Continued)	
	<u>Part II</u>	
	Test Results and Suitability	51
	<u>Part III</u>	
	Item Analysis	54
V	SUMMARY AND CONCLUSIONS	63
VI	RECOMMENDATIONS	70
	 BIBLIOGRAPHY	 73
	APPENDIX A	76
	APPENDIX B	77

LIST OF TABLES

Table		<u>Page</u>
I	Summary of Data Select-R Standardization Groups	20
II	Summary of Data Select-A Sample Stratified for Schooling by Province	20
III	Racial and Regional Origin of Experimental Sample	20
IV	Distribution of Scores on Select-R Obtained by Experimental Group	40
V	Comparison of Statistics on Select-R for Experimental and Standardization Samples	40
VI	Distribution of Scores on Select-A Obtained by Experimental Group	42
VII	Comparison of Statistics on Select-A for Experimental and Standardization Samples	42
VIII	Intercorrelations of Select-A and Select-R with Validation Criteria for Experimental Sample	50
IX	Percentage of Correct Responses on Select-R for Experimental and Standardization Samples	55
X	Proportion of Correct Responses on Select-A for Experimental and Standardization Groups	57

LIST OF FIGURES

Figure		<u>Page</u>
I	Graphical Representation of Mean Percentage of Correct Responses on Select-R Test	56
II	Graphical Representation of Mean Proportion of Correct Responses on Select-A Test	58

CHAPTER I

INTRODUCTION

General Considerations

In the history of psychological testing there have been many attempts to derive instruments measuring intelligence or ability which, to a greater or lesser degree, are free of influence from the culture in which the person taking the test was reared. Although it is generally accepted that " the range of mental capacities in all ethnic groups is much the same", most tests now in use, as Vernon points out (28), appear to discriminate against individuals who are not native to the culture or society in which the test was developed. There is even some strong argument that it is a misconception to suggest that a culture-free test can be developed (13). Nevertheless, several tests have been developed which attempt to reduce cultural bias in intelligence testing. (6, 18, 26, 29).

During the 1950s, Doctor R.A. Wendt and associates constructed a test known as the Canadian Army Select-R test for use as a preliminary screening device in the Canadian Army; the test is currently in use for that purpose. It was designed with a view to reducing cultural bias. Between 1952 and 1956 validation studies were carried out with the test and the results, in the view of Wendt, indicated that it did not discriminate between "culturally restricted" and non-restricted groups and that it measured abilities which did not vary systematically with degree of education, hence, "acculturation"

(30). A further study involving outport Newfoundlanders, considered to be culturally restricted, revealed no significant differences from the mean score of the standardization samples in the other Canadian provinces (2). No further studies have been completed with any group, using this test.

Between 1958 and 1960 another screening device, the Select-A Test, was developed by the Canadian Army to identify applicants for enrolment who would be unlikely to obtain satisfactory aptitude test scores at a later stage in processing. This test was constructed with no attempt to reduce cultural bias. Validation studies revealed significant differences in mean scores in various Canadian regions and for varying degrees of education (8). Since 1960 no further validation studies have been carried out on this test.

Although the Canadian Forces recruit throughout the country, very few young men of non-white racial origin in the Northwest Territories and Yukon Territory applied for enrolment until very recently. No evaluation of the screening efficiency of the Select-R and Select-A tests in this population had therefore been undertaken. However, studies by Doctor R.S. MacArthur in the early 1960s, using a number of intelligence tests, including some which are designed to be "culture-reduced", among Indian and Metis youth in Northern Alberta and the Northwest Territories suggest that a definite cultural bias against such persons exists in using such tests. Accordingly, it is believed that the Select-R and Select-A tests would probably show a similar bias against Indian, Eskimo and Metis youth in the

northern regions. Such a bias, if it exists, could very well render the two screening devices very limited, even virtually useless, as aids in determining the suitability of applicants for service in the Canadian Forces.

In developing and using psychological tests of ability for any purpose, whether in connection with suitability for higher academic training, assessment of capacity for learning technical skills of a higher order or selection of applicants for service with the armed forces, it is most desirable to obtain measures which are valid cross-culturally. That is, every test of ability contains some material which may be peculiar to a particular societal group and is dependent on the environmental background of the individual tested. The aim of such tests must be to assess those abilities which are not affected to any great degree by the testee's cultural background. This is particularly so in the increasingly mobile modern world where sub-groups from many cultures can be found in almost any country.

Canada, especially, is a nation whose population is very heterogeneous with respect to cultural background yet, as in the United States, all are expected to cope with training and employment based primarily on Anglo-Saxon traditions and conducted in the English language. Tests currently in use for assessment of suitability for such training and employment are essentially based on this type of background, yet there are many sub-groups, such as the Indians and Eskimo, who have not been assimilated completely into the predominantly Anglo-Saxon society. Standard tests of ability might well, therefore,

be unsuitable for use with such groups. To facilitate their assimilation, it is necessary to find and employ selection tests which are relatively free of specific cultural and environmental influences.

Recent changes in Canadian Forces recruiting procedures have now made available a sizable group of young men of non-white origin for testing the Northwest Territories. The Department of Northern Affairs requested that the Canadian Forces exercise some measure of flexibility in enrolment standards to enable more youth from the Territories to gain employment in the remainder of Canada. Because of the uniform training programme for recruits from all parts of Canada and evidence of adequate predictive validity of the screening devices currently in use, the Department of National Defence was unwilling to arbitrarily amend enrolment standards without a scientifically-proven basis for such an action.

Applicants for service in the Canadian Forces are initially processed through Edmonton if they are resident in the northern Territories. The author, engaged in post-graduate studies at University of Alberta, learned of the problem in the North and, foreseeing the theoretical and practical psychological possibilities, secured authority to carry out a study of the efficiency of the Select-R and Select-A tests in predicting the suitability of non-white applicants in the Territories for service in the Canadian Forces. Normally, only one of the two devices is used to initially screen applicants. If one can be shown to be more efficient and

unbiased as a predictive instrument, enrolment procedures could be amended accordingly.

The Problem

The purpose of this study was to determine the degree to which the Canadian Army Select-R and Select-A Tests are effective and useful as pre-enrolment screening devices when employed with a Canadian Forces applicant population considered to be "culturally-restricted". It was hypothesized that the tests, having been developed for and standardized on a non-restricted Canadian applicant population, would show a discriminative bias against a non-white and "culture-reduced" sample from the Northwest Territories. The experimental sample was tested and compared with the standardization samples upon which the tests were developed.

For a variety of reasons unconnected with this study, it was not possible to enrol any sizable number of applicants from the experimental sample and place them in military training. This precluded assessment of test scores of the experimental group against the most important criterion variable, i.e. success or failure in military training.

Accordingly, using previous studies with Select-R, Select-A and "culturally-restricted" Canadian groups as a guide, the criteria selected for assessing the effectiveness of the two tests with the experimental sample were: (1) significant similarity of mean test scores of the sample with those for the standardization samples; (2) low correlation of test scores with educational level and age; and (3) high correlation with a dichotomous, subjective assessment of each applicant's suitability for enrolment, made by the testing officer.

For interest, the inter-correlation of the two tests was also calculated. Where possible, these correlations were compared with relevant data for the standardization groups.

Further, to set the stage for additional research at a later date, a detailed item analysis was undertaken to determine those items in the tests on which the specific sub-culture samples in the study differed significantly in number of correct responses as compared with the standardization samples in previous studies.

Because the study was primarily concerned with the objective utility of the devices in a specific context - effectiveness of the tests as they are now constituted as initial screening tests for military service - no attempt was made to suggest possible changes in the format or nature of the instruments to improve them as psychometric measures of ability.

Definition of Terms Used

In this dissertation the terms "culture-free", "culture-reduced" and "culturally-restricted" have been used above but must now be functionally defined in order to determine the theoretical context in which the tests used in the study is set. Since the tests are being applied in this study to a societal group which has been isolated from the remainder of Canada by geography, travel limitations, social contacts and a radically different cultural background, it is very important to define such terms pertaining to culture. The tests used have been adopted as aids in predicting success in the military milieu with the general Canadian applicant population and are, perforce,

the instruments by which the non-white Northern groups are assessed, since no independent selection tests exist and the Forces require that all those enrolled, no matter what their background, meet the same training standards.

This last requirement suggests that any test used for the prediction of success in military service should be as free as possible of variance associated with the language, educational level and cultural background of the applicant. Little problem develops with respect to language since one of the criteria for enrolment in the Canadian Forces is that the applicant be able to speak, read and write either English or French and the screening tests are available in both those languages. The minimal educational requirement of at least Grade VI for enrolment at least partially eliminates the problem of variance associated with educational level; at worst this criterion eliminates a segment of the applicant population which would most likely be more unsophisticated in dealing with formal tests than the better educated. Thus the problem basically is reduced to one of the effect of culture on test results. Accordingly, it is desirable to use only those tests which have had some efforts to reduce cultural influences incorporated in their construction. To the degree that this effort is successful will depend the efficiency of the tests for such purposes as predicting success in military service in a country, such as Canada, which includes applicants from many cultural backgrounds. If the tests used are highly "culturally-loaded", there will be

inevitable discrimination against some groups within the applicant population; conversely, a test with little influence from the cultural background of the applicant should yield a useful and meaningful measure of potential for success in training.

However, the question of cultural influence in such tests is directly related to the theoretical problem of what is meant by "intelligence", which is what most such instruments purport to measure.

Hebb (16) has argued for two postulates of intelligence which he styles Intelligence A, or innate potentiality, and Intelligence B, which is the product of potentiality and environmental stimulus, or developed mental efficiency. In lay terms these may be described as heredity and environment, and the combination of these two. Others, including Ferguson, Skeels and Harms, and Vernon (12, 27, 28) have suggested that it is not possible to accurately delineate what proportion of intelligence is attributable to heredity and what to environment. All tests of intelligence must also take into account variance associated with other factors, such as chance, physiological defects, the physical setting in which the test is administered, the use to which results are to be put, the motivation of the subject, etc. All these can vastly affect the scores obtained. Wendt (30), in developing the Select-R test, concluded that intelligence tests are essentially sampling Intelligence B, that is, a combination of innate or inherited potentiality and environmental stimulus.

This leads to the conclusion, arrived at by Ferguson (13), that there is no such thing as a "culture-free" test nor can there be, since the skills measured by intelligence tests are acquired through social and physical environmental influences. The term "culture-free" is therefore deemed rather meaningless and will not be further used in this study.

It is, however, theoretically and practically possible to construct a test measuring the construct "intelligence" but which is "culture-reduced", that is, which can be applied cross-culturally through elimination or reduction of the influence of items which have peculiar reference to a specific societal organization. In today's world of high speed transportation and mobility of populations, this is of considerable importance, especially in cases where test results are to be used in a multi-cultural setting where all those tested are expected to meet certain standards later, as in the Canadian Forces. Cattell, Kahn, Safran and Wendt (6, 18, 26, 30) have all produced tests purporting to measure intelligence, which attempt to reduce cultural influence through the use of symbolic items of simple perceptual content. The use of non-verbal, low difficulty items thus lowers the effect of the specific culture in which the test subject was reared, though the fact that some form of verbal communication in setting up the test situation will still exercise some cultural influence. So too will level of sophistication of reasoning in the educational frame-work of the individual's native culture affect his performance on a test which attempts to

reduce cultural bias. The employment of symbolic items does tend to mitigate this influence, as does the use of low difficulty items. Using these as the criteria for a "culture-reduced" test, it is hypothesized that the Select-R Test can accurately be defined as "culture-reduced", while the Select-A Test cannot be so considered.

It is, perhaps, more difficult to adequately define the expression "culturally-restricted" with respect to a particular population. First, one must establish and define a frame of reference for the conceptualization of a non-restricted culture, i.e. not limited to a specific cultural background. In this study, the standardization samples for the Select-R and Select-A tests were drawn from the total Canadian population, permitting the arbitrary assumption that this is a non-restricted cultural group for the purpose to which the tests are put. Within the limits of the other Forces selection standards, applied prior to the administration of a screening test, to wit, sex, age, citizenship and basic education, the tests were developed and standardized on a population not otherwise restricted.

Assuming, then, that the total Canadian population is non-restricted, the population from which the experimental sample will be drawn can be regarded as "culturally-restricted". Until construction of the Distant Early Warning system in the 1950s, the Indian and Eskimo bands in the Northwest Territories were in contact with the Canadian society in the South only through missionaries, government officials, Mounted Police officers, Hudson's Bay traders, trappers and miners. Some

bands rarely or never came into contact with white persons. The last ten years have changed this somewhat but there are still many villages and encampments which see only police officers with any frequency. Centuries-old traditions of living persist, formal education in schools is often not available and many of the artifacts accepted as normal in the South are non-existent, even unknown, in the North. There is no television, few books; trains, automobiles and modern plumbing are virtually unknown in the more remote areas. Life is basically nomadic, following the caribou, seal and whale supply. The sample can thus rightly be described as coming from a culturally-restricted population, though all applicants have received some formal schooling.

At risk of introducing another term into an already confused situation, it might be argued that the experimental sample hails from a population which is "culturally-different" than that on which the tests used were standardized. This term has fewer semantic connotations and less suggestion of racial differences than the terms "culture-reduced" and "culturally-restricted". It can certainly be made abundantly clear that the non-white population of the Northern areas is different in culture to that of the rest of Canada. However, to maintain uniformity with other studies with such groups, the terms "culture-reduced" and "culturally-restricted" will be retained in this study.

CHAPTER II

SURVEY OF THE LITERATURE

History of "Culture-Reduced" Tests

The history of tests which attempt to avoid discrimination against cultural groups and environmental influences dates back to Brigham's 1923 book (3) which suggested that certain national groups have greater average intelligence than others. Following a number of other studies of culturally-restricted groups, Brigham modified his views in 1930 (4) and Estabrooks (11) found no reason to support the evidence for the doctrine of racial differences in intelligence, though he expressed doubt that valid techniques for investigating cross-cultural differences could be developed.

In 1926 Pintner developed his non-language test (25) but made some very strong assumptions regarding a more or less common civilized background of test subjects, though these are not necessarily valid assumptions. Later studies by Mead (19) and Clements (32) demonstrated that Pintner's assumptions did not hold, due to test content. Subsequent studies took the latter consideration into account (1, 5, 22) and proved to some degree that test material had an important bearing on results in testing of groups not reared in the culture within which the tests used were developed.

The study by Skeels and Harms (27) demonstrated that hereditary factors are relatively unimportant in intelligence test scores. This was followed by Hebb's assertions (16) regarding early life experience, which must be taken into

account in selection of appropriate test items, though Hess (17) showed the difficulty in selecting items involving real life and providing no cultural advantage.

A number of other studies, well reviewed by Wendt (30), indicate that "non-language intelligence tests render more accurate estimates of intelligence of groups whose experience has been different from that of the standardization population than do verbal tests". (14, 15) This has been borne out by studies involving the Cattell Culture-Free Intelligence Test (6, 7), which is based on 'g' saturation, 'a la' Spearman, Kahn's New Culture-Free Intelligence Test (18) and the Safran Culture-Reduced Intelligence Test (26).

All the foregoing tests tend to bear out Hebb's hypothesis of a distinction between innate potentiality (Intelligence A) and developed mental efficiency (Intelligence B), which is the product of potentiality and environmental stimulus. (16) In addition, while arguing that development of a truly "culture-free" test is not possible, Ferguson (12) has given a valuable insight into the construction of valid, cross-cultural, non-verbal tests with his concept that intelligence consists of over-learned skills and techniques for coping with problems which have crystallized out of the child's previous experience and which can be transferred to a wide variety of new situations. If one can find a group of items which sample these skills or techniques without regard to verbal fluency or linguistic ability, it is theoretically possible to construct a test which strongly reduces cultural bias.

It was in such a context and with such views that Wendt approached the construction of the Select-R test, which was to "be relatively free of variance associated with widely differing environmental influences." (30)

Development of the Select-R Test

The Canadian Army in the early 1950s was desirous of having a screening test which would, in Wendt's words (23), "yield scores independent of language, cultural and marked environmental differences in the population to be tested", in this case the applicants for Army service from across Canada. Since it had been proven during the Second World War that the Army can train illiterates, a non-verbal test has distinct advantages. Other criteria which such a screening device must meet are a short time limit, capable of being administered to groups and with separate answer sheets for stencil scoring. It is also desirable that it be self-administering except for total illiterates. Wendt set out to develop a test meeting these requirements which would be free of cultural bias only for the population of Canadian Army applicants and which screened out those who would be least capable of absorbing basic military training.

From a comprehensive study of the history of so-called "culture-free" tests, he concluded that construction of a test meeting these criteria is possible. (30) He based his work on two assumptions: "(1) that one must avoid items which involve relatively complex, culture-specific abilities, and (2) that it is possible to abstract, empirically or rationally, a sufficient sample of appropriate culturally-common items."

(23) He inferred that pattern analogy, paper form board and picture absurdities items of low difficulty would be appropriate.

His other inferences were that "non-verbal items would be superior to verbal items; that items requiring the identification of a non-common picture would be of about the same order of difficulty for both high and low status groups; that one might expect to sample abilities which are more obvious in verbal items even though non-verbal items are used (and) that these inferences were more likely to be borne out empirically if items of low, uniform order of difficulty were used". (30)

From previous studies conducted with applicant groups he chose three types of items: picture absurdities, figure analogies and paper form board. A fourth type of item, labelled concept formation, was devised by Wendt. These items required identification of a fifth object which did not belong in the same category as the other four pictured in each item.

The items used were amended to include five alternate responses and organized to permit the use of a separate answer sheet. Seven practice items which were purposely made as difficult were developed to illustrate each type of item.

Two experimental trials were conducted to arrive at an optimal time limit under power conditions and to assess item difficulty. A 12 minute time limit was settled upon and the test was printed with a re-arranged order of items.

The third trial, with a sample of 988 subjects, using the 12 minute time limit, was found to be representative of

the same Canadian population which Dickie (10) had sampled during the Second World War.

The test, administered under these conditions, met expectations as to item difficulty and false leads. "The distribution was such that maximum discrimination could be expected through the middle 75% of the range." (23)

Estimates of reliability, concurrent, predictive and construct validity were obtained. Reliability was assessed by the test-retest method, with 179 applicants being re-tested after a 28-day interval. The product-moment correlation obtained was $r = .758$, with mean gain on re-test of 4.54 points. Concurrent validity was investigated by comparing subject's Select-R scores with scores obtained on the Army selection device then in use (Revised Examination "M") (8) and years of schooling. The product-moment correlation for Select-R and "M" was $r = .74$, indicating reasonable concurrent validity in terms of a well proven test of intelligence. Partial r between Select-R and education with "M" held constant was $r = .021$, and that between "M" and education was $r = .384$. This indicated that Select-R measured abilities which did not vary systematically with education, hence acculturation.

A comparison of the scores of three groups of individuals assessed as "culturally-restricted" with the remainder showed that there were significant differences at the .001 level on "M" but differences for Select-R were not significant at the .05 level. The test was therefore considered to be relatively free of cultural bias.

Predictive, or criterion, validity was studied by correlation of Select-R scores with, first, a dichotomous subjective assessment of each applicant's intellectual capability or incapacity for completing military training carried out by a psychologically trained Personnel Selection Officer before enrolment of the applicant, and, second, the success of each enrolled applicant at two stages of initial training, e.g. after eight weeks (Depot training) and sixteen weeks (Recruit training). The criterion for those enrolled and placed in training was successful completion of training to the stages mentioned, as assessed by officers and non-commissioned officers at training depots, according to long-established standards of proficiency in such subjects as drill, first aid, handling of small arms, deportment, adjustment to discipline and elementary field tactics. The bi-serial correlation between Select-R score and the trainable/untrainable dichotomy was $\sqrt{.835}$. Correlation coefficients between Select-R scores and enrolled applicants who had completed Depot training and Recruit training were $\sqrt{.321}$ and $\sqrt{.213}$ respectively, very much lower than the correlation of the Select-R scores with the subjective assessment of suitability.

Construct validity was investigated in a factorial study of the Select-R and the eight tests in the Revised Examination "M" battery, then used in the Canadian Army as a selection test but since succeeded by other tests. Four principal factors were extracted: perceptual speed, verbal, mechanical and extrapolation. Wendt (30) found the factor analysis to show that the

Select-R had appreciable loadings in all the factors which are correlated with the several tests contained in the "M" battery. The inter-correlations of the Select-R and the eight sub-tests of the "M" ranged from .609 with the Verbal Analogies sub-test to .368 on the Vocabulary sub-test, on a sample of 946 persons tested with both devices.

The Select-R loading on the perceptual speed factor, after three rotations, was .611. All eight sub-tests of the "M" had loadings significantly different from zero on this factor. Five of the sub-tests of the "M" had significant loadings on the mechanical factor while the Select-R loading on this factor was significant but relatively low (.223). Only three of the "M" sub-tests had significant loadings on the extrapolation factor. The Select-R loading on this factor was .449, higher than any of the significant "M" sub-tests.

The verbal loadings are of especial interest. The Select-R loading on this factor was .186, which was a significant variation from zero. The three non-verbal sub-tests on the "M" Test had verbal factor loadings which were not significant, while the other five sub-tests, which have definite verbal content, had much higher loadings on this factor than the Select-R. Wendt concluded that the low but significant loading on Select-R for this factor was probably explained by the fact that "the instructions, which are given verbally, at the beginning of the test, are subject to variable retention" (30). This conclusion is by no means satisfactory

as an explanation for the significant verbal loading but no other readily offers itself.

In general, the Select-R had significant loadings on all the factors extracted from the "M" battery but none of the individual sub-tests of the battery had the same generality of significant loadings on all factors. Since the Select-R is measuring the same abilities as those sampled by the "M" battery, Wendt concluded that "the Select-R is a valid measure of intelligence as measured by the "M" test" (30). The practical significance of this finding is that the Select-R is both quicker and easier to administer than the "M" battery, yet it yields a sound indication of intelligence, as this is measured by the "M" test.

The mean scores and standard deviations of the sample of 988 applicants and a subsequent sample of 1364 applicants are given in Table I. The test showed considerable evidence of consistency in operation between different groups of applicants since neither the mean nor standard deviation is significantly different between the two samples. Wendt construed this as evidence of test reliability in the sense that "a cutting score will eliminate the same proportion of individuals in successive groups" which he feels, perhaps with some justification, is what one truly wants to know about group tests, as applied in the military setting. (30)

From the data cited, Wendt concluded that the test was both a reliable and valid instrument for pre-enrolment screening and that it was relatively free of cultural and environmental

Table I

Summary of Mean Scores and Standard Deviations
Select-R Standardization Groups (1954 and 1956)

Statistics	1954 Sample	1956 Sample
Mean	39.35	39.04
SD	9.29	9.55
N	988	1364

Table II

Test Results of Select-A Sample Stratified for
Schooling by Province (June 1959 - December 1960)

Statistics	Select-A	"M" Test
Mean	40.3	134.8
SD	14.3	23.1
N	1000	1000
r	.79	

Table III

Racial Origin and Regional Origin of Experimental
Sample of Indians, Eskimo and Metis (1965)

Region	Indian	Eskimo	Metis	Total
Mackenzie Delta and Arctic	18	12	7	37
Yellowknife and Fort Simpson	10	0	3	13
Total	28	12	10	50

bias. Accordingly he published his data and the Select-R Test was taken into use by the Canadian Army as a screening device in 1956. Nevertheless, the very low correlation of Select-R scores with subsequent success in training is such that it can be seriously doubted whether or not the Select-R is a valid indicator of potential for success in military training. Nor is the evidence of reliability of the Select-R Test conclusive. Though it correlates quite highly with the "M" test and the dichotomous assessment of suitability of applicants by Personnel Selection Officers, it did not prove very valid as a predictor of success in training, as the low correlations in this respect indicate. Thus, while perfectly acceptable as a non-verbal and relatively "culture-reduced" screening test, it cannot be inferred to be a valid and reliable test of training ability in the usually accepted sense of these criteria, on the basis of the obtained data.

In fact, the findings reported by Wendt concerning the Select-R, the "M" test and the assessment of applicants by Personnel Selection Officers suggest that none of these methods are very effective in predicting success in military training. Though there is a good measure of correlation between the three, the absurdly low correlation between the Select-R and later success of those enrolled indicates that it might be well to abandon current procedures and either find other selection tests or enrol every applicant who applies. This is, however, not economically feasible in terms of monetary expense, the staff to handle such large numbers of would-be

servicemen and the physical plant necessary to accommodate them. The search therefore continues for more acceptable and valid measures of ability to succeed in military training. The development of the Select-A Test was one step in this direction.

Development of the Select-A Test

By 1958 it had been decided that a short and easily administered test of intelligence was required for use by Army recruiters to screen applicants for enrolment who would be unlikely to meet any one of the wide range of standards expressed as scores on the Revised Examination "M", which would be more closely correlated with the "M" test than the Select-R and which included verbal content. The "M" was then in effect as the main selection device at Personnel Depots but, since it takes over one hour to administer, is not re-usable, requires extensive scoring procedures, yields multiple scores and must be interpreted by highly-trained personnel, it was considered desirable to have a short, verbal test which could be quickly administered and scored by relatively untrained recruiting personnel in the field. The Select-R meets most of these criteria but is non-verbal and the correlation with Revised Examination "M" was assessed as too low.

Major A.B. Laver and Lieutenant L.M. Kendall at Army Headquarters in Ottawa undertook to develop a test meeting the criteria above. The Select-A Test was the results of their efforts. Results of the research are contained in two Canadian Army publications (8, 24) which are on restricted distribution due to security regulations.

Items for the Select-A were adapted from items in one form of the Canadian Army Classification Test, which was an omnibus test of intelligence with multiple choice and open-end items, developed during the Second World War. The latter test, in a wartime study, correlated $r = .81$ with the "M" test. It contained fourteen practice items and seventy-five test items.

To prepare Form I of the Select-A all 89 items of the CAC Test were changed to multiple-choice form with five alternate responses. The four wrong answers for each item were so framed as to be plausible false leads. A French form, which was a direct translation of the English version, was prepared. The items probably measure at least three factors of ability: (1) a general reasoning factor; (2) a verbal comprehension factor; and (3) an education factor; according to the authors.

Form I was administered to 108 English-speaking applicants at three Personnel Depots and 126 French-speaking applicants at two other Personnel Depots in April and May 1958, just prior to administration of the "M" test. The mean score and standard deviation of the English-speaking sample were 48.06 and 14.36 respectively. From this trial a maximum time limit of 30 minutes was set and item difficulties were calculated.

On the basis of the first trial administration, items were reordered on the basis of difficulty values, from easiest to most difficult, eight of the fourteen practice items were eliminated, to shorten administration time, and the test

was reprinted as Form II. This form was administered to all applicants processed at all Personnel Depots in June 1958. Number of items completed at the end of varying time periods was recorded and "M" scores were obtained at the same time. From the data thus obtained it was ascertained that a time limit of 22 minutes was optimal for both language versions, though there was no statistically significant difference between correlation coefficients between Select-A and "M" scores for different time limits.

On the basis of the findings to that time, the Select-A Form II was taken into use in the Canadian Army in June 1959, replacing the Select-R Test. All used answer sheets were forwarded to Army Headquarters, along with other information on each applicant tested. By the end of 1960 more than 9000 answer sheets were available. Parameters were calculated for the English language sample based on 7217 cases. The mean and standard deviation for this sample were 34.5 and 14.6, respectively.

From the large English Language sample, a smaller sample of 1000, stratified for schooling by province, was extracted. Using this sample, statistics are as given in Table II. The correlation coefficient between Select-A and "M" is high, suggesting that the Select-A may be used for selection and classification decisions instead of the "M" test when testing time and scoring ease are relatively important considerations.

The analysis of variance technique was used to test the significance of the differences between the mean Select-A

scores of men with different amounts of schooling from each of the ten provinces. There was a marked tendency for the Select-A score to increase as the last school grade completed increased. In addition evidence of regional differences at most grade levels was noted, with the country being divided on this basis into three groups: Newfoundland, the other Maritime provinces, and the remainder of Canada. A tentative conclusion was drawn that the regional differences were attributable more to local employment conditions and educational opportunities than to true differences in level of ability, since proportionally there are more Army applicants from the Atlantic provinces than the rest of Canada.

To test the hypothesis that there are no differences between Select-A scores because of age, the analysis of variance technique was applied to a sample of 1043 applicants from Ontario with Grade VIII schooling, tested in English. No significant differences were found in ten one-year age groups from ages $15\frac{1}{2}$ to $25\frac{1}{2}$.

Reliability of the Select-A Form II was calculated by the Kuder-Richardson method for a sample of 555 applicants tested under speed-power conditions. The coefficient of internal consistency reliability was .925 with an SE of measurement of ± 5.56 .

This evidence of concurrent validity and reliability led to the adoption of the Select-A Test in lieu of the Select-R as the Canadian Army's preliminary screening device from 1960 onward. In 1965 the Select-A also replaced the Revised Examination "M" as the selection and classification test for

all applicants for service in the Canadian Forces except female and officer applicants. Pending further standardization studies, arbitrary cutting scores for enrolment were established for Navy, Army and Air Force respectively.

With this test, as with the Select-R, doubts can be expressed as to the validity of the device in predicting success in military training. The test authors themselves acknowledge this but indicate that "it is known that "M" test scores are related to training success, and the high correlation between Select-A scores and "M" test scores suggests that it can be reasonably assumed, in the absence of information to the contrary, that Select-A scores are also related to training success". (24) They further indicate that decisions based on both scores should be more valid than those based only on the "M" score and contend that the use of the Select-A by recruiters will produce a curtailment in the "M" test distribution at Personnel Depots. By reason of the high correlation between the "M" test and the Select-A, they estimate that the percentage reduction in errors of prediction will be of the order of 59% but do not explain why they arrived at this figure. They do, however, argue that this is evidence of the concurrent validity of the Select-A.

Even granting that their conservative (sic) estimate is correct, the figure arrived at is not exceptionally high. It cannot be taken as strong evidence of the concurrent validity of the device. While there is unquestionably some value in administering the Select-A test at recruiting

stations or in the field, thereby reducing the number of men of very low ability who would otherwise be sent to Personnel Depots only to be rejected there on the "M" test score, this economy says nothing about the validity of the Select-A as a test of mental ability or its efficiency in predicting success in training. Nothing has been done with the test since its adoption to overcome this deficiency or otherwise prove its validity. This being so, the device can only be regarded as a useful aid in screening out applicants of very low ability before they reach Personnel Depots.

Studies on Culturally-Restricted Canadian Groups

As mentioned above, only one study has been undertaken with the Select-R Test on a population deemed to be culturally-restricted, that of Beach in Newfoundland in 1956.

(2). No such studies whatever have been attempted with the Select-A Test.

In point of fact, there have been extremely few studies of the applicability of intelligence tests to those groups in Canada which may be considered culturally-trestricted, especially those in the Northern territories. MacArthur, and students working with him, have produced the only serious study of this segment of the Canadian population (20, 21, 31), insofar as can be found in the literature.

Beach took a sample of 83 Newfoundlanders, 64 from Outport villages, who are generally considered to have been reared in a culture-reduced environment, and administered the Select-R, Otis Higher Form A and Progressive Matrices Revised Order 1956.

The mean Select-R score of the group was 37.55, standard deviation 8.57, which was not significantly different from the mean of the 988 standardization sample. The mean Otis I.Q. obtained in the group was 89.5. Assuming the mean Otis I.Q. of the general population to be 100 (since no data exists for a similar sample of Canadian adults), the Newfoundland sample scored significantly lower than the population mean. Correlation of Select-R with Otis was .72, Select-R and Progressive Matrices .57 and Otis with Progressive Matrices .62. When the Otis was partialled out the correlation between Progressive Matrices and Select-R (.23) was not significantly different from zero at the one percent level of confidence.

The evidence obtained by Beach tends to confirm Wendt's contention that the Select-R test is relatively free of variance arising from culturally-restricted background, compared with such standard verbal tests as the Otis.

In 1962 MacArthur reported to the Canadian Psychological Association on a study of the intellectual ability of Metis children in Northern Alberta, later publishing a paper on the study (21). He found that the Progressive Matrices, Cattell Test of g Scale 2 and Safran Culture-Reduced Intelligence Test showed less bias against the Metis than commonly-used group intelligence tests and moderate correlation with school achievement. He later extended his study to an Indian-Metis sample at Fort Simpson, N.W.T. (20).

The sample consisted of 155 Indian and Metis of both sexes attending school in Grades I through IX. Culture-reduced tests used included Progressive Matrices, Cattell

Test of g Scale 2, Safran Culture-Reduced Intelligence Test and Lorge-Thorndike Non-Verbal Intelligence Tests. Among the standard intelligence tests used were the California Short-Form Test of Mental Maturity, the Detroit Beginning First-Grade Intelligence Test and the Otis Quick Scoring Mental Ability Tests.

Using the same criteria as with the Northern Alberta study, minimum cultural bias and moderate correlation with concurrent school achievement, the mean scores of the sample on Progressive Matrices, Catell Test of g and Safran and, to some degree, California Test of Mental Ability Non-Language were significantly different from the California Total, Detroit and, in some cases, Lorge-Thorndike means. The same tests correlated to a much lesser degree with the California Achievement total than did the verbal tests employed, including Detroit and Otis.

From the data, MacArthur concluded that Progressive Matrices and Safran satisfied the criteria of construct validity, that is, less cultural bias than conventional intelligence tests, and concurrent validity, namely, moderate correlation with school achievement.

On the basis of the foregoing, it was concluded that the Select-R and Select-A tests would probably show cultural bias against the Indian, Eskimo and Metis of the Northwest Territories because they were not constructed using standardization groups from that area and, in the case of the Select-A, correlated so highly with level of education, hence, acculturation. With more applicants for service coming from the

Territories, a study of the usefulness of the screening devices in the culturally-restricted group appeared desirable, leading to this dissertation - the first of its kind in the Canadian Forces and preceded only by MacArthur's investigation at Fort Simpson.

CHAPTER III

PROCEDURE

This chapter will deal with a description of the measuring devices used in this study; the method of their administration and scoring; the selection of the experimental group and description of the standardization groups from previous studies with the devices, used for comparison purposes; and a general discussion of the method of determining the results.

Description of the Select-R Test

One of the tests used in this study is the Canadian Army Select-R Test, which consists of 65 items composed of several kinds of simple perceptual problems, i.e. geometrical figures or simple pictorial representations or stylized drawings. Seven practice problems are included. Items are of a multiple-choice type involving a simple discrimination, requiring the subject to select either the picture that is different from four others, one that is left over when the other four are paired, the one that is wrong, or, the one that is right. The items are incorporated in a test booklet and answers are given on a separate answer sheet. All parts of each item in the test booklet are labelled A, B, C, D or E. The answer sheet is composed of columns of the same letters opposite a number corresponding to the item number in the booklet. The answer sheet is marked by crossing out the appropriate letter for each number. The test booklet

contains carefully worded general instructions which are amplified by instructions for the completion of the sample problems. The test can be self-administered to literate subjects; for illiterates instructions are read to the subject by the administrator. A French language version has been prepared through direct translation. Scoring is carried out by means of a cardboard overlay with circles punched out for the correct response. The sum of the correct responses yields the raw score. Answer sheets are checked for multiple responses to items; items for which there are more than one response are not included in the raw score.

The test was developed and standardized on the applicants for service in the Canadian Army from across Canada during January and February 1954. Neither education nor previous intelligence tests scores were used as standards for selection. The test was found to have the following characteristics: (1) useful criterion validity, (2) satisfactory concurrent validity, (3) satisfactory construct validity, and (4) satisfactory reliability. A complete description of the test can be found in Major R.A. Wendt's paper (30) and Personnel Selection Service Technical Report No. 5 (23).

Description of the Select-A Test

The second device used in this study is the Canadian Army Select-A Test, an objective group test of general ability, made up of 75 multiple choice problems of various kinds, arranged in order of increasing difficulty. The problems include items concerned with ability to invent

solutions to problems (such as arithmetic reasoning and verbal reasoning), ability to understand words (synonyms and antonyms) and the ability to see the relationship between two things or ideas and use the relationship to find other things or ideas (verbal and number relationships). Six practice problems are included. A French version was prepared through direct translation. The items are incorporated in a text booklet and answers are given on a separate answer sheet. All parts of each item in the booklet are labelled a, b, c, d or e. The answer sheet is composed of columns of the same letters. Marking of the answer sheet by the subject and scoring with a marked plastic scoring key are carried out as with the Select-R Test. Instructions are contained in the test booklet and administration requires the subject to be literate. The sum of the correct responses yields the subject's raw score, with items for which there are multiple responses on the answer sheets being scores as wrong.

The test was developed and standardized on the applicants for enrolment in the Canadian Army across Canada between June 1959 and December 1959, except that officer applicants were not included in the sample. No attempt was made to use education or other test scores as selection standards. A further sample obtained from June 1959 to March 1960 was stratified for schooling by province. The test was found to have good concurrent validity and satisfactory test-retest reliability. A complete description of the test is contained in Personnel Selection Service Technical Report No. 8 (24)

and the Canadian Army Manual for Personnel Officers (8).

Test Administration and Scoring

Test administration for the experimental group was carried out by the writer and Flight Lieutenant John Thomas, RCAF, a Canadian Forces Recruiting Officer, trained in administration of screening tests, in Yellowknife and Inuvik, N.W.T. in June and November 1965. Select-R Test was administered first, followed immediately by administration of Select-A Test. Subsequent to test administration, each subject was interviewed by the test administrator to obtain background information regarding the subject's family history, education, occupational background, reason for seeking enrolment and preference for component of the Services, i.e. Navy, Army or Air Force. On the basis of this information alone, the interviewing officer then made a dichotomous subjective judgement of the applicant's suitability for enrolment in the Canadian Forces if medically fit. To ensure uniformity of questioning concerning the subject's background, a list of questions prepared by the writer beforehand was used for each interview. This list is given in Appendix A.

Following interview and subjective assessment of suitability, the Select-R and Select-A tests were scored by the interviewing officer. All scores were recorded on the answer sheets, which were collected by the writer, along with other pertinent information. Scoring of answer sheets was later re-checked by the writer and two non-commissioned

officers, trained in test administration and scoring, employed in the writer's office.

Use of the dichotomous subjective assessment of suitability for enrolment by a trained Personnel Selection Officer as a basis for comparison with test scores was dictated by the facts that this is one of the main selection methods in the Canadian Forces, had been similarly employed in the standardization of the Select-R Test and was deemed an acceptable substitute for actual enrolment and training of all the subjects in the study, which latter was precluded for a number of reasons. Such an assessment is admittedly weak as one of the criteria employed for comparison purposes but is considered as acceptable because it is established policy in the Forces and will remain as such for some considerable time to come. It resembles the interviewing method much employed in civilian industry which usually combines an assessment by trained personnel staff with results of psychological tests in determining whether or not a prospective employee should be hired. Its primary fault is that it is a purely subjective procedure and, as shown in standardization of the Select-R, it does not necessarily correlate well with later success in training.

Selection of the Experimental Group

The subjects in this study comprising the experimental group consisted of 50 Indian, Eskimo and Metis males, permanently resident in the Northwest Territories, who had applied for or expressed an interest in serving with the

Canadian Forces. To satisfy military selection requirements, all must have completed Grade VI and all were between the ages of 15 and 30. None had previously been tested for military service. Origin of the applicants, by race, is shown in Table III. Data given is based on information volunteered by the subjects and confirmed by reference to records of the Royal Canadian Mounted Police, Department of Northern Affairs and National Resources and the schools in the Northwest Territories attended by the subjects.

All subjects had taken at least six years of schooling in English and were able to converse readily in English, though their native language included French, Eskimo, Dogrib and Slave Indian dialects.

Composition of the Select-R Standardization Group

The subjects used by Wendt in his study for standardization of the Select-R test consisted of 988 male applicants for service in the Canadian Army (Regular) in January and February 1954. The minimum education of the sample, dictated by enrolment standards, was Grade VI and the age range was from 16 to 29 inclusive. The mean educational level of the sample was 8.1 completed grades of schooling. No intelligence test scores were taken into account in selection of the group. Tests were administered at thirteen Personnel Depots across Canada. The sample consisted of young men from every type of cultural background in Canada, without discrimination as to race, creed, country of origin or location of upbringing (rural or urban).

Of the total of 988 subjects, 128 were categorized subsequently by personnel officers as having had restricted cultural experience during the first ten years of life, as compared with the generality of Canadians. Of this group, 103 had been reared in isolated communities in Canada and 25 had spent their early years in Western Europe (exclusive of the British Isles). (30)

The later study by Beach in Newfoundland involved 83 subjects, 64 of rural and 19 of urban origin, with a mean education of Grade X and educational range of Grades III through XI. Of this group, 69 were medical assistants and 14 nurses-in-training. (2)

Composition of the Select-A Standardization Group

By the end of 1960 the Select-A Test has been in use for more than eighteen months and more than 9000 answer sheets were available. Similar to the Select-R standardization group, this sample consisted of male Army applicants from all across Canada, who were tested without regard for previously administered intelligence tests. The educational range of the sample was from Grade VI to Grade XII. The age range, due to changed enrolment standards, was 16 to 25 years of age inclusive. Standardization data given in this dissertation for the Select-A is based on the 7217 English-speaking applicants in the total sample. No effort was made to make any other discrimination, such as race, cultural background or area of origin (rural or urban), in the sample. (8)

Method of Determining the Results

Mean scores and standard deviations on Select-R and Select-A Tests were obtained from results of the experimental group and compared with published data for the standardization samples by the t-test for means. Product-moment correlations were also obtained between Select-R and education, Select-A and education, Select-R and age, Select-A and age, Select-R and subjective assessment of suitability for service and Select-A and subjective assessment of suitability. For interest, the correlation between Select-R and Select-A scores for the experimental group was also computed. Finally, answer sheets were analyzed individually by item count to determine those items which appear to discriminate against the experimental sample and, therefore, the cultural group from which it was drawn. The results are discussed in the next chapter.

CHAPTER IV

RESULTS

This chapter will consist of three parts. Part one will outline the findings for the experimental group and compare them with the parameters previously published for the standardization groups. The bearing this comparison has on the hypothesis that the tests show a cultural bias against the experimental group will be discussed. In addition, this section will discuss the results obtained with the experimental group when correlated with educational achievement and age, as well as the inter-correlation of the two tests used.

Part two will examine the relation of the test results with the experimental group to the subjective assessment of suitability for military service and the implications of these results.

Part three will outline the findings obtained from detailed item analysis of the tests used with the experimental groups, and the effect these findings have on the employment of the tests in the "culturally-restricted" group studied.

Part I

Select-R Scores and Comparison of Groups

The experimental group had a mean score of 44.30 on Select-R, with standard deviation of 7.80. Table IV shows the range of Select-R scores and the data which were used to obtain the mean and standard deviation. Table V gives the statistics for the standardization group of 988 used by Wendt and compares the standardization sample with the

Table IV

Distribution of Scores on Select-R
Obtained by Experimental Group (N - 50)

Interval	Frequency
61-65	1
56-60	2
51-55	6
46-50	15
41-45	12
36-40	7
31-35	5
26-30	1
21-25	1
Total	50

Table V

Comparison of Statistics on Select-R
for Experimental and Standardization Samples

Statistics	Experimental Sample	Standardization Sample
Mean	44.3	39.35
SD	7.8	9.29
N	50	988
t	4.66 #	

- Significant at .001 level

experimental group.

From these tables it is clear that the experimental group obtained a mean score considerably above the mean for the standardization group. The t test¹ of the significance of the difference between the means of the two samples yielded a value of t of 4.66, which is significant at the .001 level of confidence. Since the mean scores of the group classed by Wendt as "culturally-handicapped" in his study (38.13) and that of the later sample of 1364 applicants (Table I) of 39.04 are lower than the total standardization sample, it is obvious that the difference between the means of those groups and the experimental sample is also significant.

The standard deviation of the experimental group is somewhat below that for the standardization sample but this may be due to the relative homogeneity of the experimental sample and the small number of subjects in the study, rather than to any significant difference in the groups. The significance of the difference in means will be discussed more fully below.

1 The formula for determining values of t is as follows: (9)

$$t = \frac{D_{M1-M2}}{\sigma_{\overline{D_M}}} \qquad \sigma_{\overline{D_M}} = \sqrt{\frac{(\sigma_1)^2}{N_1} + \frac{(\sigma_2)^2}{N_2}}$$

Table VI

Distribution of Scores on Select-A
Obtained by Experimental Group (N - 50)

Interval	Frequency
56-60	1
51-55	0
46-50	4
41-45	5
36-40	5
31-35	6
26-30	8
21-25	10
16-20	4
11-15	3
6-10	3
Total	50

Table VII

Comparison of Statistics on Select-A
for Experimental and Standardization Samples

Statistics	Experimental Sample	Standardization Sample
Mean	29.2	34.5
SD	11.7	14.6
N	50	7217
t	3.19#	

- significant at .001 level

Select-A Scores and Comparison of Groups

The mean Select-A score of the experimental group was 29.2, with a standard deviation of 11.7. Both these values are considerably below that for the standardization sample (mean 34.5, standard deviation 14.6). Statistics on Select-A for the experimental group are given in Table VI. Table VII shows the statistics for the 1959-60 standardization sample of 7217 applicants and the comparison of that group with the experimental sample. A t test of the significance of the difference between the mean scores of the two samples yielded a value of t of 3.19, which is significant at the .001 level of confidence.

As with the standard deviation of the experimental sample on Select-R scores, the standard deviation of the group on Select-A is considerably below that of the standardization sample but is likely accounted for by the same factors which pertained to the Select-R: more homogeneous group and small number of subjects. The significance of the difference in means will be discussed below.

On the basis of these simple statistics alone, it appears that the Select-R test fulfills Wendt's desire to construct a test relatively free of variance associated with cultural and environmental differences, insofar as the non-white population of applicants for service in the Northwest Territories is concerned. The method of selecting the samples was identical except that the experimental group expressly consisted of applicants of other than Caucasian

race. The testing procedure was not changed.

However, the possibility exists that the experimental sample was basically different in some respects from the standardization sample and that these differences account for the significantly different mean scores of the two groups. For example, it has often been suggested that the Canadian Forces applicant population contains a high proportion of men who have been unable to make adequate adjustment to civilian jobs through lack of education, basic intellectual inferiority and emotional maladjustment. Certainly, most applicants for service seek enrolment because of dissatisfaction with their success as civilians, are primarily in the unskilled or semi-skilled labour categories, are usually unemployed at the time of applying for service and have a mean education of approximately Grade VIII. Thus, the Forces are scarcely attracting applicants from the total spectrum of the eligible population (8). It is therefore conceivable that the sample on which the Select-R was standardized was significantly lower in ability than the experimental group. The latter group consisted largely of young men who had attended or were attending Federal Government schools which are highly institutionalized and operated on a hostel system, in which students from outlying settlements attend a central school in the larger Northern centres. Further the employment situation in the Territories is such that many natives, no matter how well educated and intelligent, cannot find jobs locally and must search for positions outside the Territories. Taking these factors into

account, the experimental sample could well consist of men who are truly superior in ability on the average as compared with the group of applicants on which the Select-R was standardized.

Nevertheless, even if such is the case, the data obtained in this study suggests that the Select-R Test is equally as applicable as a screening device for service in the Canadian Forces with the non-white applicants from the Territories as it is with the applicant population in the remainder of Canada. It can therefore justifiably be regarded as "culture-fair" or "culture-reduced", in that there is no evidence that it displays a bias predicated on different environmental background with respect to the population from which the experimental sample was drawn.

On the other hand, the significantly lower mean score of the experimental group on the Select-A test suggests that this device has limitations with respect to its use as a pre-enrolment screening test with respect to the non-white applicant population from the Territories, as compared with the Select-R. The criteria for selection of the standardization samples for both tests were the same. Though composed of different subjects, the Select-R and Select-A standardization samples were drawn from an applicant population meeting the same general requirements as to age, education and background. The two standardization samples must therefore be assumed to be comparable. The experimental sample was selected in the same manner and employing the same criteria as with the standardization samples, except

for racial and geographical specifications. Results of the experimental group on the Select-R test, as reported above, suggest that the group was superior to the standardization sample with respect to those abilities measured by the test. If the Select-R and Select-A standardization groups are assumed to be comparable, it is therefore true by definition that the experimental sample should achieve approximately the same mean score on Select-A test, at least, since they did better on Select-R.

The significantly lower mean score of the experimental group on Select-A must then be accounted for by some factor or factors which do not appear, or appear to a lesser degree, on Select-R. The main difference is that one is non-verbal and claimed to be relatively free of bias due to acculturation while the other is highly verbal and highly correlated with education. If the conclusions arrived at by Beach and MacArthur (2, 20, 21) are correct, it is probable that the lower mean score of the experimental sample on Select-A reflects the effects of the restricted cultural environment in which the sample was reared and the consequent lesser capacity of the group for coping with tests containing verbal material and items relating to skills and artifacts relating to their native culture. The practical application of the finding is more fully discussed below.

Inter-correlations and Correlations

Inter-correlations between the tests used in the study

and correlations of the tests used with education and age were computed using a product-moment coefficient of correlation formula¹, while correlations between test scores and subjective assessment of suitability were calculated using a bi-serial coefficient formula². (9) Table VIII shows the intercorrelations for Select-A and Select-R tests with education, age, assessment of suitability and each other for the experimental sample, with appropriate data for standardization groups, where applicable.

The correlation between Select-R and Select-A scores for the experimental group is relatively low (.49), suggesting that these tests are sampling different skills, as might be expected since the one is highly verbal while the other is not and employs different types of items. The standard error of estimate in predicting the Select-A score from the Select-R score was in the order of 10.18, indicating that Select-R score should not be arbitrarily used as an estimate of Select-A score for the group from which the experimental sample was drawn. This conclusion is further borne out by the small percentage of the variance (24%) which the tests share in common for the experimental sample.

1

$$r = \frac{\overline{UV} - \bar{U}\bar{V}}{\sigma_U \sigma_V}$$

2

$$r_{bis} = \frac{M_P - M_T}{\sigma_T} \left(\frac{p}{z} \right)$$

These results further demonstrate that, all other factors being equal, Select-R test scores should be used in the selection of recruits from the "culturally-restricted" population, because it has already been shown that it seems to have less cultural bias against the group than the Select-A test.

As the correlations between the tests and education show, the Select-A scores are much more closely correlated with educational level than are Select-R scores, with both the experimental and the standardization samples. There is no significant difference in the coefficients of correlation between Select-A and education for the two groups but the Select-R correlates considerably less with education for the experimental sample than for the standardization group. That education is likely a much more important factor in determining performance on Select-A is also indicated by the variances calculated from correlation coefficients. Some 34.8% of the difference in Select-A scores can be attributed to education, while only 7.8% of the difference in Select-R scores can be ascribed to this factor, for the experimental group.

It has already been demonstrated that educational opportunities for non-white residents of the Territories are somewhat more restricted than for the average Canadian youth in other parts of the country. This being so, it is preferable to use a test, such as Select-R as a screening device with the Northern applicants since it is less influenced

by conditions and level of schooling than the Select-A test, or other highly verbal testing devices. Both the Select-R and Select-A tests are available for use as screening devices with the population from which the experimental sample was drawn but the use of the Select-R is to be desired in preference to the Select-A because of the former's relative dissociation from the influence of formal schooling.

Both tests used yielded a negative correlation with age, suggesting that as the age of the applicant increases, score on Select-A and Select-R is likely to decrease. This is consistent with previous results for Canadian Army applicant samples on Revised Examination "M" and Select-A tests. (8) Because of the very small number of applicants over 20 years of age in the experimental sample, it is not possible to draw any specific conclusions regarding the negative correlations obtained but it is probable that the same operative factors are in effect with this study group as with previous studies. That is, it is inferred that "decrease in the effective - or remembered - educational skills with increase in time out of schools is a factor..... It is also possible that older applicant groups include relatively more men of inferior ability; men of superior ability have less difficulty finding satisfactory civilian jobs." (8)

This does not, of course, explain entirely the relatively small correlation between Select-R and education since level of schooling does not seem to be a major factor

Table VIII

Inter-correlations of Select-A and Select-R with
Validation Criteria for Experimental Sample (N=50)

	Experimental		Standardization	
	Select-A	Select-R	Select-A	Select-R
Education	.59	.28	.659	.406
Age	-.225	-.403	---	---
Suitability/ Unsuitability	.671	.789	---	.835
Select-A	---	.49	---	---

NOTE: Insufficient data were available in the published literature on the development of the tests to show a full correlational matrix.

affecting scores on the device. However, the contention that more men of inferior ability apply from the older age groups is plausible in that only one out of the nine applicants over 20 years of age, in the experimental sample, obtained a Select-R score of more than 44, which was the sample mean. The mean education of this sub-group is almost the same as that of the entire sample (7.4 completed years of schooling against 7.62 years), therefore some other factor was likely responsible for the very low correlation of test scores with education.

Poor quality appears to be the most likely factor, since only one of the nine men over 20 years of age was recommended as suitable for enrolment by the Recruiting Officer. In all other cases, a combination of low level of education, lack of application to employment and criminal record led to a recommendation of unsuitable for service. This tends to substantiate the contention that more men of inferior ability seek enrolment at the upper age levels because they have been unable to adapt satisfactorily in civilian life; hence, lower test scores can be expected of this group, as was the case in this study.

Part II

The use of psychometric tests as military screening devices can be justified in the final analysis only if they effectively eliminate from the applicant stream those who are unlikely to succeed in basic training. Such tests are,

perforce, administered by personnel with relatively little training and experience in selection, so it is necessary that they, rather than the administrator, effectively discriminate between those who will succeed in later stages of enrolment processing and those who will be rejected on the basis of other psychological tests and an assessment by a professionally-qualified Personnel Selection Officer.

Both the Select-R and Select-A correlate highly with the "M" test, which has been the primary selection test in the Army for many years and has been found to be valid as an intelligence test. However, it is essential that the newer devices also relate well with the assessment by Personnel Selection Officers, if they are to achieve their purpose. Results of this study indicate that the Select-R and Select-A tests are useful as substitutes for the more sophisticated evaluation with other psychological tests and by trained Personnel Selection Officers, when employed by the normal recruiter in the field, who is without specialized knowledge of tests and selection procedures. This is demonstrated by the high correlation of the test scores with the dichotomous assessment of suitability for service by the interviewing officer responsible for the final decision on enrolment of the applicant. Though this assessment in itself is not too reliable, it has proven more efficient than operating with tests alone in the Canadian Army since 1940. (9)

On the whole, the Select-R scores correlated more highly with the subjective assessment than the Select-A,

with the Northern sample. The correlation of .789 is not demonstrably different from the correlation of .835 obtained with the original standardization sample. On the other hand, the correlation of Select-A scores with assessment of suitability is significantly lower than that of the Select-R for the sample studied. No comparable data are available for the Select-A standardization group.

It has already been shown that the Select-R has other qualities which commend it for use with the "Culturally-restricted" population studied; the correlation with subjective assessment of suitability further substantiates this point. On the other hand, the Select-A has sufficiently high correlation with the subjective assessment to warrant its use if the Select-R were not available. However, it would not apparently be as effective for this purpose with the non-white applicant population in the Territories as the Select-R test, likely due to its high verbal content and the influence of educational level on Select-A scores.

As a matter of interest, four of the experimental sample were further processed for enrolment and three were enrolled in the RCAF; the fourth applicant failed to achieve a satisfactory score on the Revised Examination "M", which was subsequently taken out of use. This youth has therefore been recalled for further processing. Three applicants, including the youth later rejected, obtained both Select-R and Select-A scores above the mean for the entire sample. The three who were enrolled are still in service and doing well in training.

Part III

In constructing the Select-R test, Wendt tried to select items which were culturally common for all Canadians who might have to take the test so as to reduce the bias associated with environmental influence. To the degree that he was successful in this attempt depends the efficiency of the test, in large measure. As this study is directly concerned with the validity of the Select-R test in a culturally-restricted population, an item analysis was conducted to determine those items which display some bias against the individual applicant from a restricted background.

The Select-A has been found to be influenced very greatly by educational level and, therefore, acculturation. In conducting this study, it was thus deemed fruitful to carry out an item analysis with this test to ascertain the extent to which acculturation influenced responses with the experimental group.

Table IX gives the data obtained from the sample tested with the Select-R test and compares these data with those for the group of 988 on which the test was standardized. Figure I is a graphical representation of the data contained in Table IX.

Table X contains the data obtained for Select-A test from the experimental sample and compares these data with those for an Ontario sample of 108 subjects who were tested on Form I of the device and whose scores were analysed item by item. Figure II is a graphical representation of the data contained in Table X.

Table IX

Percentage of Correct Responses on Select-R for
Experimental (N-50) and Standardization (N-988) Samples

Item	Experimental Sample	Standard- dization Sample	Item	Experimental Sample	Standard- dization Sample
8	100%	100%	41	80%	74%
9	100	100	42	88	73
10	198	100	43	90	67
11	998	95	44	58	68
12	100	94	45	76	55
13	100	95	46	78	59
14	94	93	47	76	55
15	98	95	48	80	61
16	98	90	49	80	49
17	92	89	50	74	53
18	86	90	51	70	60
19	100	96	52	58	46
20	100	94	53	70	50
21	92	88	54	48	39
22	88	84	55	62	48
23	98	87	56	34	35
24	88	79	57	48	43
25	90	85	58	44	33
26	80	87	59	32	40
27	92	82	60	52	38
28	82	78	61	46	30
29	64	78	62	4	12
30	88	68	63	28	20
31	98	83	64	26	19
32	94	87	65	30	16
33	90	75	66	6	10
34	68	74	67	14	9
35	84	67	68	6	5
36	78	71	69	8	44
37	78	83	70	16	4
38	70	75	71	6	4
39	80	70	72	6	2
40	82	69			

FIGURE I

Mean Percentage of Correct Responses on Select-R Test for
Experimental Sample (N - 50) and Standardization Sample (N - 988)

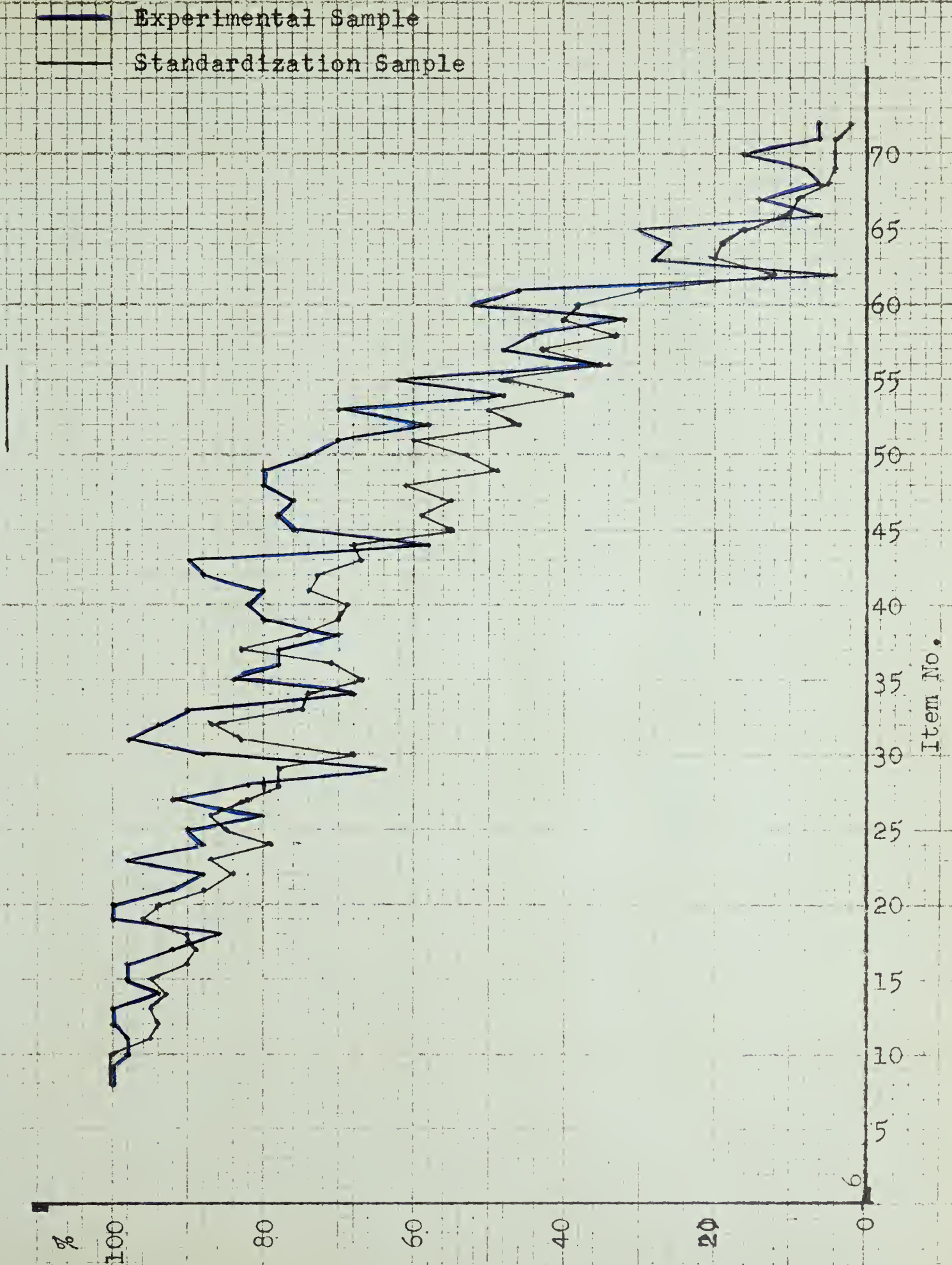


Table X

Proportion of Correct Responses on Select-A for
Experimental (N = 49)# and Ontario Form I Groups (N = 108)##

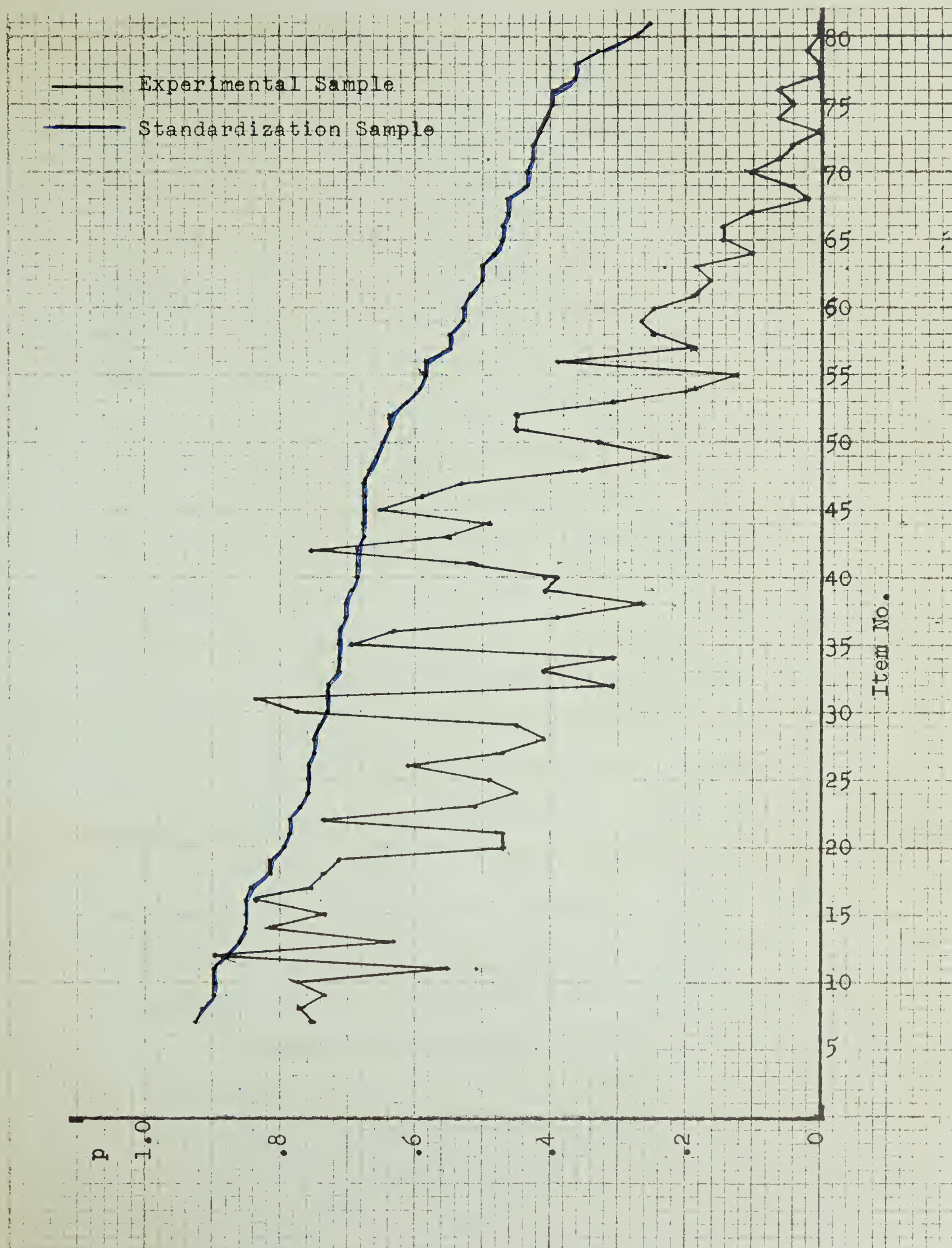
Item	Experimental Sample	Standard- dization Sample	Item	Experimental Sample	Standard- dization Sample
7	.755	.926	45	.653	.676
8	.775	.917	46	.592	.676
9	.735	.898	47	.531	.676
10	.775	.898	48	.347	.667
11	.551	.898	49	.224	.657
12	.898	.880	50	.327	.648
13	.633	.861	51	.449	.639
14	.816	.852	52	.449	.639
15	.735	.852	53	.306	.611
16	.837	.852	54	.184	.593
17	.755	.843	55	.122	.583
18	.735	.815	56	.388	.583
19	.714	.815	57	.184	.546
20	.469	.796	58	.245	.546
21	.469	.787	59	.265	.528
22	.735	.787	60	.245	.528
23	.510	.769	61	.184	.519
24	.449	.759	62	.163	.500
25	.490	.759	63	.184	.500
26	.612	.759	64	.102	.481
27	.469	.750	65	.143	.472
28	.408	.750	66	.143	.472
29	.449	.741	67	.102	.463
30	.775	.731	68	.020	.463
31	.837	.731	69	.041	.435
32	.306	.731	70	.102	.435
33	.408	.713	71	.061	.426
34	.306	.713	72	.041	.426
35	.694	.713	73	.000	.417
36	.633	.713	74	.061	.407
37	.388	.704	75	.041	.398
38	.265	.704	76	.061	.398
39	.408	.694	77	.000	.361
40	.388	.685	78	.000	.361
41	.510	.685	79	.020	.324
42	.755	.685	80	.000	.269
43	.551	.676	81	.000	.250
44	.490	.676	82		

- one test answer sheet was misplaced during processing and could not be located, hence N = 49.

- data based on Select-A Form I sample of 108 English-speaking applicants tested in April and May of 1958. Mean and SD of Sample were 48.06 and 14.36 respectively.

FIGURE II

Mean Proportion of Correct Responses on Select-A Test for
Experimental Sample (N-49) and Ontario Form I Sample (N-108)



From these data it can be seen that only on a very few items did the experimental sample obtain mean results significantly lower than the standardization group on the Select-R test. The situation was reversed with the Select-A test and the experimental group was significantly close to or better than the standardization sample on only a relatively small number of items. Appendix B shows the number of each item of both tests in this regard and describes the item in the case of Select-R test or records it verbatim in the case of the Select-A test.

In general, the experimental group had no difficulty with items on the Select-R, scoring as high as or higher than the standardization group. On the six items listed in Appendix B, however, the percentage of correct responses of the sample being studied was more than six percent lower than the group used for standardization, ranging as high as fifteen percent lower. Four of the items - 26, 29, 34 and 62 - clearly include content with cultural connotations and a fifth - 44 - contains artifacts (hammer and nail) which might be construed as having specific cultural content. The sixth - 59 - is a paper form board item, a type of problem which might also be considered as dependent to some degree on education or cultural background. Still, on many other items with pictures or drawings which might be assessed as oriented towards the culture in which the test was developed, the experimental sample had no difficulty as a whole. Presumably the content of such items must be

considered to be culturally common to both the experimental and standardization groups. Assuming this to be so, the difference for item 59 could be attributable to chance.

The data obtained from item analysis of the Select-R does not provide enough evidence to support a revision of the test for the non-white applicant population in the North. The experimental sample demonstrated that, on the whole, a group which is culturally-restricted can perform as consistently on the test as the group with which it was standardized. The few items which appear to discriminate against the experimental sample do not do so sufficiently to justify changing the items, thereby requiring additional standardization. This is especially the case in that the experimental sample achieved an overall mean score significantly greater than that for the standardization sample. In addition, the culturally-restricted group showed a capacity for working adequately with all types of item contained in the test. This suggests that the non-verbal, pictorial or figural type of item is well suited for use in any test which attempts to avoid cultural bias.

On the Select-A the experimental sample obtained a higher percentage of correct responses than the standardization group on only two items and were less than six percent lower on only eight other items. In general, the percentage of correct responses was considerably lower for the experimental group, ranging as much as 36% lower. All three types of item, dealing with general reasoning,

verbal comprehension and education, are included both in the grouping of those on which the experimental group were better than and lower than the standardization sample. No specific type of item was therefore more difficult or easier for the experimental sample.

The very large difference in mean scores of the Form I sample of 108 subjects and the Form II sample of 7217 likely has an effect on the degree of difference of correct responses, when the experimental sample is considered. The test authors suggest that sampling inadequacy accounts for the great difference in mean scores (24) and they may have, by chance, obtained a Form I sample which was truly higher in ability than the applicant population as a whole. Still, the only effect this large difference between the samples on the two forms has in considering the difference between the experimental sample in this study and the sample for Form I ($N = 108$, mean = 48.06) is that such differences as exist in number of correct responses are accentuated.

The interesting, and pertinent, fact is that on 48 of the items the experimental sample achieved a mean score more than 25% lower than that of the Form I standardization sample, clearly indicating that the Select-A is much more difficult for an applicant population which is "culturally-restricted". This follows the regional patterning reported by the test authors, which showed that as educational opportunities improve across Canada, Select-A mean scores tend to rise. From the standardization data and this study,

it was concluded that the Select-A Test is somewhat discriminatory against the Indian/Eskimo/Metis applicant population from the Northwest Territories. The high correlation with education and the use of verbal items which pertain to a culture relatively alien to that of the group sampled in this study indicate that the test is not as acceptable a screening test for military service to a degree even remotely approaching the usefulness of the Select-R with that specific population.

CHAPTER V

SUMMARY AND CONCLUSIONS

This study was concerned with determining the usefulness or effectiveness of two Canadian Army tests as screening devices with a "culturally-restricted" service applicant population. These tests - The Select-R and Select-A - are in use in the Canadian Forces at present to assist recruiting personnel in identifying applicants who are unlikely to succeed in military training. The "culturally-restricted" experimental sample was drawn from male Indian, Eskimo and Metis applicants in the Northwest Territories. To meet service requirements the applicants had to have completed Grade VI education and be between 15 and 29 years of age. No other criteria of pre-selection were applied.

The two tests were administered to the experimental sample by the writer and a Canadian Forces Recruiting Officer. Results were then analyzed and compared with data published for the groups on which the tests had been standardized in the remainder of Canada. The standardization groups consisted of samples of Army applicants who met the same age and educational requirements as the experimental group.

It was hypothesized that the tests, which were developed for and standardized on a non-restricted population, would show a discriminative bias against the "culturally-restricted" experimental sample from the Northwest Territories. The criteria for assessing the effectiveness of the tests with

the experimental sample were significant similarity of mean test scores of the sample with those for the standardization samples; low correlation with educational level and age; and high correlation with a dichotomous, subjective assessment of the applicant's suitability for enrolment made by the testing officer. These correlations were compared, where possible, with relevant data for the standardization groups. For interest, intercorrelation of the two tests was also calculated.

Finally, a detailed item analysis of the two tests was conducted to determine the items in the tests on which the specific "culture-restricted" sample used in the study differed significantly in the number of correct responses as compared with the standardization groups.

Comparison of the mean scores of the experimental and standardization samples on Select-R test showed that the experimental group scored significantly higher than the mean for the standardization sample, the difference being significant at the .001 level of confidence.

Quite the reverse was true of the Select-A test. The experimental group obtained a mean score significantly lower than that obtained by the standardization group, the difference again being significant at the .001 level.

Standard deviation of the experimental sample on both tests was somewhat smaller than that for the groups used in standardizing the devices, probably due to the relative homogeneity of the experimental sample and the small number

of subjects (50 in that sample.

For interest, Select-R and Select-A scores for the experimental group were correlated with each other. The relatively low correlation (.49) suggests that the two tests are sampling very different skills. This is logical in that the Select-R is a non-verbal test employing pictorial and figural items while the Select-A is a highly verbal test which is greatly influenced by educational level. From the rather low intercorrelation of the tests, it was deduced that scores on the one should not be used to predict scores on the other and that it is preferable to use the Select-R test with "culturally-restricted" populations who are adversely influenced by tests with high verbal content.

Correlation of the test scores of the experimental group with education showed that the Select-A is much more highly associated with educational level than the Select-R. This is also the case with the standardization groups. It is concluded that the Select-R is therefore a better device for initial screening for military service since it is more independent of formal education and conditions of schooling, hence, also the effects of culture and prior environment.

Both tests yielded a negative correlation with age, scores tending to decrease as the age of the applicant increases. This is consistent with previous findings with Canadian Army selection tests but some further evidence was found which suggested that the negative correlations were related to the fairly low over-all quality of those applicants over 20 years of age in the experimental sample. This was

particularly the case with respect to the high negative correlation of Select-R and age ($-.406$). The contention that more men of inferior ability seek enrolment at upper age levels was supported by the relative test scores.

The results of correlation of Select-R and Select-A scores of the experimental group with a dichotomous subjective assessment of suitability for enrolment indicated that both tests are useful as substitutes for a full-scale interview and evaluation by a professional Personnel Selection Officer, a process which is not practicable in every location where potential recruits are found. It was not possible to enrol any sizable proportion of the sample and place them in training to obtain the most acceptable criterion on which to assess the effectiveness of the screening tests. Though a subjective assessment by officers charged with recruiting and selection, including psychologically-trained Personnel Selection Officers, is not as acceptable a criterion as success in training, it is still one of the established criteria of selection and useful as an alternative in assessing the effectiveness of the tests, particularly when the results of military training are unavailable.

Again, the correlation of Select-R with the criterion of suitability for enrolment ($.789$) was considerably higher than that of the Select-A for the experimental sample but both correlations are sufficiently high to argue that the two tests are acceptable for use as screening devices in the "culturally-restricted" population studied, in the same

sense that they are useful in the general Canadian applicant population. The obtained data do, however, suggest that the Select-R is the more desirable and efficient instrument in this regard.

Detailed item analysis of test results from the experimental group revealed that no particular type of item on either test was specifically easier or more difficult for the sample, though there were considerable variations with respect to individual items when the experimental and standardization samples were compared. The Select-R contained six items which the experimental group found more difficult but, on the test as a whole, they achieved equally good or better results than the standardization group. Most of the significantly more difficult items contained culture-specific matter which probably accounts for the lower number of correct responses by the study sample. The Select-A was more difficult for the experimental sample by a considerable margin when compared with one of the groups used in standardization, with variations in percentage of correct response as great as 36% between the groups on some items.

From the item analysis, it was concluded that revision of the Select-R test is not necessary in order to use the test effectively in screening of non-white applicants from the Far North and that the non-verbal, pictorial or figural type of items are preferable to those which include verbal content. Thus, the Select-A did not prove as useful a

screening test as the Select-R in the "culturally-restricted" sample used in the study.

In summary, the obtained data partially proved the hypothesis that the two tests used in the study display a discriminatory bias against the non-white applicant population from the Northwest Territories, who were assessed as "culture-restricted". Though both tests are acceptable as substitutes for a more comprehensive assessment of suitability for enrolment, when employed by relatively unskilled recruiters in the field, the Select-R shows less bias against the population from which the experimental sample was drawn than does the Select-A test, probably due primarily to the non-berval content of the former. Accordingly, it is concluded that the Select-R should be employed as the primary and initial screening test with this population in preference to the Select-A test.

A partial summary of these results, dealing with the comparison of mean scores of the experimental and standardization groups and the correlations of test results with education and assessment of suitability for service, was forwarded, on request, to Canadian Forces Headquarters in Ottawa in March 1966, with a recommendation that the Select-R test replace the Select-A test as the initial screening device for the non-white applicant population in the Northwest Territories and Yukon Territory. This recommendation was accepted and appropriate instructions issued to

the responsible recruiting staff in May 1966. This study thus achieved its practical aim of effecting an improvement in Canadian Forces selection procedures.

CHAPTER VI

RECOMMENDATIONS

Though this study has presented some valuable and useful evidence in support of the use of the Select-R test with the "culture-reduced" population sampled, it is desirable to extend the study in several ways to further establish the effectiveness of both the Select-R and Select-A tests with such groups.

Specifically, it is most desirable to obtain larger groups of applicants of non-white origin from the Northern territories of Canada for testing because the sample used in the present study was very small and not too well-balanced between Indians, Eskimo and Metis. Of needs, only those young men who volunteer to serve in the Forces or are willing to take the tests can be sampled and the number of these from a total human population of less than 25,000 in the Territories is quite restricted. However, over a period of several years, such a sample should become available and the author may have the opportunity to further the present study.

Because of the use to which the tests employed is put, that is, screening for suitability for military training, it is deemed essential that more applicants from the "culture-reduced" population sampled be enrolled in the Canadian Forces and given training up to and including skilled tradesman and senior non-commissioned officer status. Failing this, the gathering of a suitable group of such

applicants for basic military training in situ, i.e. within the Territories, or as a special unit outside the Territories, is imperative. Only when one of these alternatives, preferably the first, is done can a proper empirical assessment of the validity of service screening devices be carried out. Certain technical considerations, including costs, lack of suitable facilities and more urgent tasks for the available training cadres, render it impracticable to carry out such a program at the moment. Still, it is hoped that the results obtained in this study will serve to hasten the day when further research can be carried out under conditions which will enable a better assessment of the predictive validity of the tests through actual training of a sizable number of recruits from the "culture-reduced" population of the Territories in a military setting.

Inasmuch as all three studies which have used the Select-R with "culture-reduced" groups - those of Wendt (30), Beach (2) and the present study - have shown it to be relatively unbiased against such groups in comparison with standard intelligence tests, it is recommended that further research be undertaken with other groups in Canada and abroad which can be similarly characterized as culturally limited. The studies so far undertaken have involved from 50 to 128 persons, primarily males. There are many Canadian sub-populations which qualify as "culture-reduced", including Indians in all parts of the country, recent

immigrants from the so-called "under-developed" countries, the Outport Newfoundlanders who Beach studied, etc. In addition, the possibilities outside Canada are virtually unlimited. As the Select-R test need not be restricted to use with Canadian Forces applicants, it would also be useful to standardize the test on females. With sufficient research along these lines, it could be determined conclusively whether or not Wendt was correct in his assumptions and whether revision of the Select-R is necessary.

Insofar as the Forces applicant population is concerned, the same recommendations with respect to Select-A apply regarding "culture-reduced" groups.

BIBLIOGRAPHY

1. Altus, W.D., A note on group differences in intelligence and the type of test employed, J. consult. Psychol., 1948, 12, p. 194-195.
2. Beach, H.D., unpublished letter to Wendt, R.A., May 1957.
3. Brigham, C.C., A Study of American Intelligence, Princeton, Princeton University Press, 1923.
4. -----, Intelligence tests of immigrant groups, Psychol. Rev., 1930, 37, p. 158-165.
5. Carlson, H.B. and Henderson, N., The intelligence of American children of Mexican parentage, J. abnorm. soc. Psychol., 1950, 45, p. 544-551.
6. Cattell, R.B., A culture-free intelligence test, I, J. educ. Psychol., 1940, 31, p. 161-179
7. -----, Feingold, S.N. and Saranson, S.B., A culture-free intelligence test, II. Evaluation of cultural influence on test performance, J. educ. Psychol., 1941, 32, p. 81-100.
8. Department of National Defence, Army, Canadian Army Manual for Personnel Officers, (1960), Vol. I, Ottawa, Queens Printer, 1960.
9. Department of National Defence, Army, Canadian Army Manual for Personnel Officers, (1960), Vol. II, Ottawa, Queens Printer, 1960.
10. Dickie, R.D., A Statistical analysis of the characteristics of a 3000 sample of Canadian Army Personnel, World War II, ORG Report No. 5, Ottawa, Queens Printer, 1952.
11. Estabrooks, G.H., That question of racial inferiority, Amer. Anthropol., 1928, 30, p. 470-475.
12. Ferguson, G.A., On learning and human ability, Canad. J. Psychol., 1954, 8, p. 95-112.
13. -----, On transfer and the abilities of man, Canad. J. Psychol., 1956, 10, p. 121-131.
14. Garth, T.R., Elson, T.H. and Morton, M.M., The administration of non-language intelligence tests to Mexicans, J. abnorm. soc. Psychol., 1936, 31, p. 53-58

15. Garth, T.R. and Smith, O.D., The Performance of full-blood Indians on language and non-language intelligence tests, J. abnorm. soc. Psychol., 1937, 32, p. 376-381.
16. Hebb, D.O., The Organization of Behaviour, New York, Wiley, 1949.
17. Hess, R.D., Controlling culture influence in mental testing. An experimental test, J. educ. Res., 1955 49, p. 53-58.
18. Kahn, T.C., A new "culture-free" intelligence test, Psychol. Rep., 1960, 6, p. 53-58.
19. Mead, M., Group intelligence tests and linguistic difficulties among Italian children, Sch. and Soc. 1927, 25, p. 465-468.
20. MacArthur, R.S., Assessing the Intellectual Ability of Indian and Metis Pupils at Fort Simpson, N.W.T., Ottawa, Department of Northern Affairs and National Resources, 1962.
21. MacArthur, R.S., The intellectual ability of Metis pupils at Faust, Alberta, in Card, B.Y. et al., The Metis in Alberta Society, Edmonton, University of Alberta, 1963.
22. Nissen, H.W., Machover, S. and Kinder, E.F., A study of Performance tests given to a group of native African negro children, Brit. J. Psycho., 1935, 25, p. 308-355.
23. Personnel Selection Service, Technical Report No. 5, Ottawa, Queens Printer, 1958.
24. Personnel Selection Service, Technical Report No. 6, Ottawa, Queens Printer, 1959.
25. Pintner, R., An empirical view of intelligence, J. educ. Psychol., 1926, 17, p. 608-616.
26. Safran, C., Safran Culture-Reduced Intelligence Test, Calgary, Calgary Public School Board, 1960.
27. Skeels, H.M. and Harms, I., Children with inferior social histories: their mental development in adoptive homes, J. genet. Psychol., 1948, 72, p. 283-294.
28. Vernon, P.E., Environmental Influences and Mental Development, address presented to Faculty of Education, University of Alberta, Edmonton, April 5, 1965.
29. Wendt, R.A., The Select-R Test, in Geldard, F. (Ed.), Defence Psychology, London, Pergamon, 1962.

30. Wendt, R.A., The Development of the Select-R Test, unpublished PhD. thesis, Carleton University, 1958.
31. West, L.W. and MacArthur, R.S., An evaluation of selected intelligence tests for two samples of Metis and Indian children, Alta. Journal Ed. Res., X, 1, 1964.
32. Clements, F., Notes on the construction of mental tests for American Indians, J. soc. Psychol., 1930, 1, p. 542-548.

APPENDIX A

QUESTIONNAIRE USED BY RECRUITING OFFICERS IN STUDY1. Education

- a. Last grade successfully completed.
- b. Grades failed and why.
- c. Age on starting and leaving school.
- d. Type of school (government, church, etc.).
- e. Location of school (town, village).
- f. Number of classrooms in school.
- g. Approximate number of students in school.
- h. Marks in each subject in last grade completed (if available).

2. Occupation

- a. Type of employment held (fishing, logging, etc.).
- b. Title of position in each job (helper, roughneck, etc.).
- c. Duration of employment in each job.
- d. Wages earned in each job.
- e. Reason for leaving each job.
- f. Length of time unemployed since leaving school.

3. Family

- a. Location of family during childhood of applicant.
- b. Occupation of father.
- c. Number of brothers and sisters.
- d. Whether or not applicant lived with family while at school.
- e. Racial or national extraction.
- f. Languages spoken at home other than English.
- g. Languages applicant can speak, read and/or write.

4. Recreation

- a. Spare time activities (sports, hobbies, reading, etc.).

5. Conduct

- a. Amount of truancy from school.
- b. Arrests, charges, convictions by police or courts.
- c. Length of time in civil custody.
- d. Extent of use of alcohol.

6. Motivation

- a. Reason(s) for applying for service.
- b. Service in which interested (RNC, Army, RCAF) and why.
- c. Type of employment desired in service (mechanic, infantryman, etc.).

APPENDIX B

ITEMS ON SELECT-R ON WHICH EXPERIMENTAL GROUP
OBTAINED MUCH LOWER SCORES THAN STANDARDIZATION
SAMPLE (Explanation of non-verbal items)

<u>Item No.</u>	<u>Content</u>
26.	Selection of incorrect animal among pig, horse, mouse, cow and sheep.
29.	Selection of wrong picture from infant in crib, being bathed by mother, floating in tub on lake, playing with blocks and playing with kitten.
34.	Selection of incorrect article among pair of shoes, suit jacket, pair of spectacles, hat and pair of gloves.
44.	Selection of correct picture among five showing hammer being used to pound nail into board.
59.	Selection of correct paper form to match broken form.
62.	Selection of incorrect tool among saw, hammer, pipe wrench, scissors, and pliers.

ITEMS ON SELECT-A ON WHICH EXPERIMENTAL GROUP
OBTAINED EQUAL OR HIGHER SCORES THAN STANDARDIZATION
GROUP (Verbatim)

12. What number comes next in the following series?
17, 16, 15, 14, 13, 12
(a) 13 (b) 14 (c) 25 (d) 10 (e) 11
14. The opposite of believe is
(a) doubt (b) expect (c) trust (d) hope
(e) argue
16. Horse is to hair as fish is to
(a) swim (b) hook (c) water (d) scales
(e) ocean

<u>Item No.</u>	<u>Content</u>
22.	What number comes next in the following series? 5, 15, 6, 16, 7, 17, 8, 18 (a) 11 (b) 9 (c) 16 (d) 10 (e) 19
30.	If 4 rifles cost \$240, what will 9 rifles cost? (a) \$450 (b) \$290 (c) \$540 (d) \$360 (e) \$480
31.	The opposite of <u>defend</u> is (a) support (b) attend (c) attach (d) amend (e) intend
35.	If it is 360 yards around a square parade ground how many yards is it along each side? (a) 180 (b) 90 (c) 60 (d) 120 (e) 80
42.	Which word does not belong in this list? (a) stream (b) creek (c) river (d) brook (e) pond
45.	What number comes next in the following series? 192, 96, 48, 24 (a) 6 (b) 16 (c) 12 (d) 48 (e) 0

B29865